

Input file Flh14273new; Output File Flh14273tra

Sequence length 1743

TCCGGACTAGTTCTAGACCGCTGCGGGCCGCCAGGCGCCGGGA	M S P E C A R A A	9
ATG TCC CCT GAA TGC GCG CGG GCA GCG		27
G D A P L R S L E Q A N R T R F P F F S		29
GGC GAC GCG CCC TTG CGC AGC CTG GAG CAA GCC AAC CGC ACC CGC TTT CCC TTC TTC TCC		87
D V K G D H R L V L A A V E T T V L V L		49
GAC GTC AAG GGC GAC CAC CGG CTG GTG CTG GCC GCG GTG GAG ACA ACC GTG CTG GTG CTC		147
I F A V S L L G N V C A L V L V A R R R		69
ATC TTT GCA GTG TCG CTG CTG GGC AAC GTG TGC GCC CTG GTG CTG GTG GCG CGC CGA CGA		207
R R G A T A C L V L N L F C A D L L F I		89
CGC CGC GGC GCG ACT GCC TGC CTG GTA CTC AAC CTC TTC TGC GCG GAC CTG CTC TTC ATC		267
S A I P L V L A V R W T E A W L L G P V		109
AGC GCT ATC CCT CTG GTG CTG GCC GTG CGC TGG ACT GAG GCC TGG CTG CTG GGC CCC GTT		327
A C H L L F Y V M T L S G S V T I L T L		129
GCC TGC CAC CTG CTC TTC TAC GTG ATG ACC CTG AGC GGC AGC GTC ACC ATC CTC ACG CTG		387
A A V S L E R M V C I V H L Q R G V R G		149
GCC GCG GTC AGC CTG GAG CGC ATG GTG TGC ATC GTG CAC CTG CAG CGC GGC GTG CGG GGT		447
P G R R A R A V L L A L I W G Y S A V A		169
CCT GGG CGG CGG GCG CGG GCA GTG CTG CTG GCG CTC ATC TGG GGC TAT TCG GCG GTC GCC		507
A L P L C V F F R V V P Q R L P G A D Q		189
GCT CTG CCT CTC TGC GTC TTC TTT CGA GTC GTC CCG CAA CGG CTC CCC GGC GCC GAC CAG		567
E I S I C T L I W P T I P G E I S W D V		209
GAA ATT TCG ATT TGC ACA CTG ATT TGG CCC ACC ATT CCT GGA GAG ATC TCG TGG GAT GTC		627
S F V T L N F L V P G L V I V I S Y S K		229
TCT TTT GTT ACT TTG AAC TTC TTG GTG CCA GGA CTG GTC ATT GTG ATC AGT TAC TCC AAA		687
I L Q I T K A S R K R L T V S L A Y S E		249
ATT TTA CAG ATC ACA AAG GCA TCA AGG AAG AGG CTC ACG GTA AGC CTG GCC TAC TCG GAG		747
S H Q I R V S Q Q D F R L F R T L F L L		269
AGC CAC CAG ATC CGC GTG TCC CAG CAG GAC TTC CGG CTC TTC CGC ACC CTC TTC CTC CTC		807
M V S F F I M W S P I I I T I L L I L I		289
ATG GTC TCC TTC TTC ATC ATG TGG AGC CCC ATC ATC ATC ACC ATC CTC CTC ATC CTG ATC		867
Q N F K Q D L V I W P S L F F W V V A F		309
CAG AAC TTC AAG CAA GAC CTG GTC ATC TGG CCG TCC CTC TTC TTC TGG GTG GTG GCC TTC		927

Fig 1

T F A N S A L N P I L Y N M T L C R N E 329
ACA TTT GCT AAT TCA GGC CTA AAC CCC ATC CTC TAC AAC ATG ACA CTG TGC AGG AAT GAG 987

W K K I F C C F W F P E K G A I L T D T 349
TGG AAG AAA ATT TTT TGC TGC TTC TGG TTC CCA GAA AAG GGA GCC ATT TTA ACA GAC ACA 1047

S V K R N D L S I I S G * 362
TCT GTC AAA AGA AAT GAC TTG TCG ATT ATT TCT GGC TAA 1086

TTTTTCTTTATAGCCGAGTTTCTCACACCTGGCGAGCTGTGGCATGCTTTTAAACAGAGTTCA^{TTT}CCAGTACCTCCA

TCAGTGCA^{CCCT}GCTTTAAGAAAATGAACCTATGCAAAATAGACATCCACAGCGTCGGTAAATTAAGGGGTGATCACCAA

GTTTCATAATATTTTCCCTTTATAAAAGGATTTGTTGGCCAGGTGCAGTGGTTCATGCCTGTAATCCCAGCAGTTTGGG

AGGCTGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAACCTGACCAACATGGTGAGACCCCGTCTCTACTA

AAAATAAAAAAAAAAATTAGCTGGGAGTGGTGGTGGGCACCTGTAATCCTAGCTACTTGGGAGGCTGAACCAGGAGAAT

CTCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCGTGCCATTGCACTCCAACCAGGGCAACAAGAGTGAAAC

TCCATCTTAAAAAAAAAAAAAAAAAGATTTGTTATGGGTTCTTTTAAATGTGA^{ACT}TTTTTTAGTGTGTTTGTAATATG

ATCAAATTTAATAAATATTTATTTATGACTGTTT^{CAG}CAAAAAAAAAAAAAAAAAAAGGGCGG

RTA01/2057957v1

10075987 " 021302

FIG 1 (CONT)

Sequence	Description	Score	E-value	N
7tm_1	PF00001 7 transmembrane receptor (rhodopsin)	119.9	4.7e-37	1

Parsed for domains:

Sequence	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
7tm_1	1/1	57	321	..	1 259 []	119.9	4.7e-37

Alignments of top-scoring domains:

7tm_1: domain 1 of 1, from 57 to 321: score 119.9, E = 4.7e-37

```

*->GNLLVilvilrtkkkrtptnifilNLAVADLLflltltppwalyylvg
GN+ ++++++r +rr +t +++NL ADLLf + p++ ++ ++
Flh14273, 57 ONVCAVLVAR-RRRRGATACLVNLFCADLLFISAIPVLAVR-WT 101

gsedWpFGsalCKlvtaldvnmYaSillLtaISiDRYLAivhPlryrrr
e W++G++ C+l+ ++++++ + il+L+a S++R + Iv l+ +r
Flh14273, 102 --EAWLLGFVACHLLFYVMFLSGSVTILTLAAVSLERMVCIV-HLQRGVR 148

rtsprrAkvvillvwlalallalPpllfawvktveegngtlnvavvcli
+r +v++l+W ++++++lP +f+ v + ++ ++ ++ +C++
Flh14273, 149 GPGRRARAVLLALIWGYSAVAALPLCVFFRVVPQR-LPG--ADQEISICTL 196

dfpeestasvstwlrayvllstlvqFlilPlilvilvcYtrIlrtlr.....
+p++++ ++s+ ++ ++ Fl+P lvi++ Y+ Il + + +++++
Flh14273, 197 IWPTIPG-----EISWDVSFVTLNFLVPGLVIVISYSKILQITKasrkx 240

.....kaaktllvvvvvFvLCWlPyfiwilldtlc
+ + +++++ + +++++ ++ +tl++++v F++ W P i++li +
Flh14273, 241 lcvalayseshqirvsqgdfrLFRTLFLLMVSEFFIMNSPIIITILLIIQ 290

.lsiimsstCelervlptallvclwLayvNsclNPiY<-*
++ + + p +++++ + +++++Ns+LNpi+Y
Flh14273, 291 nFK-----QDLVWPSLFFWVVAFTFANSALNPILY 321

```

//

FIG 2

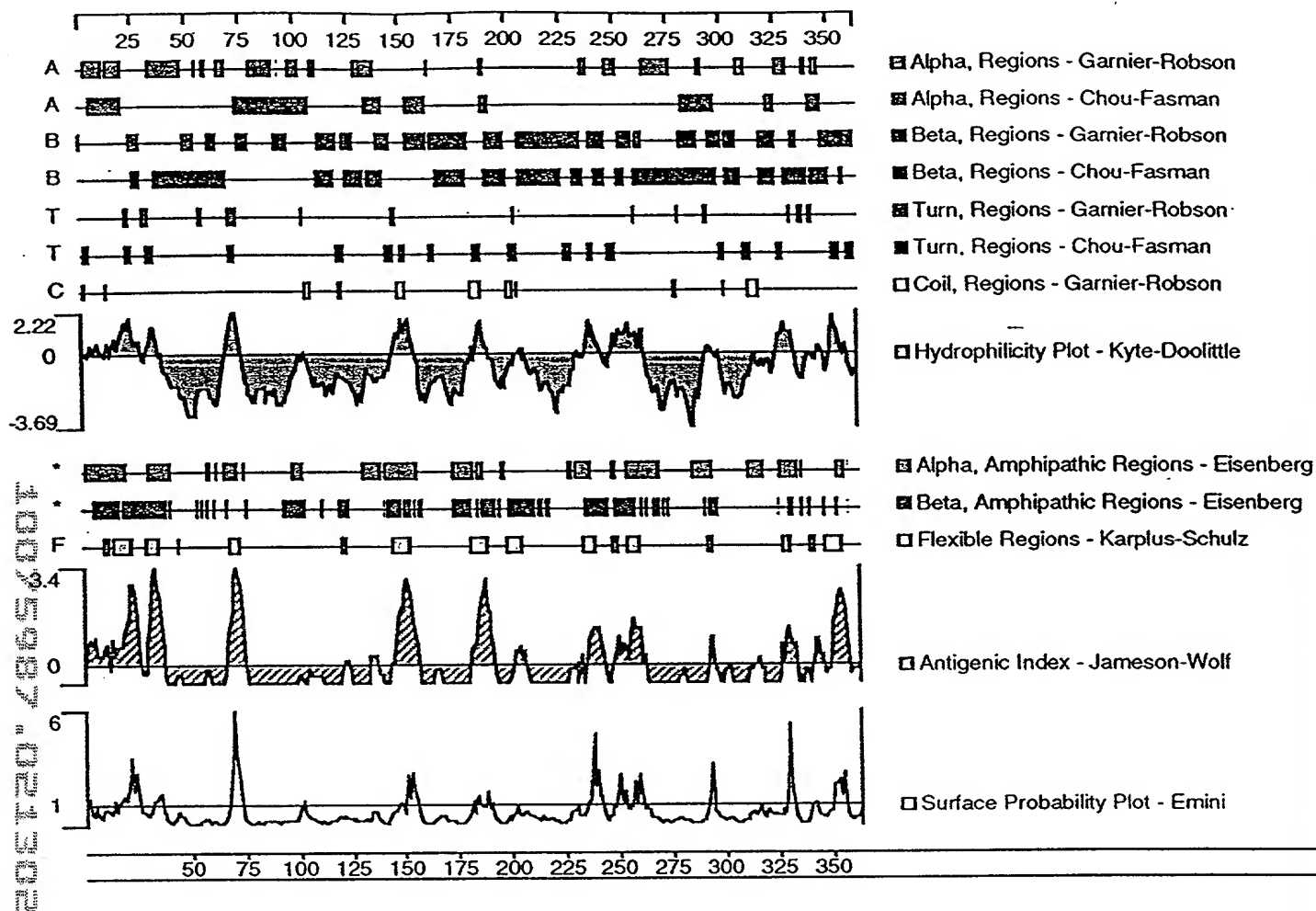
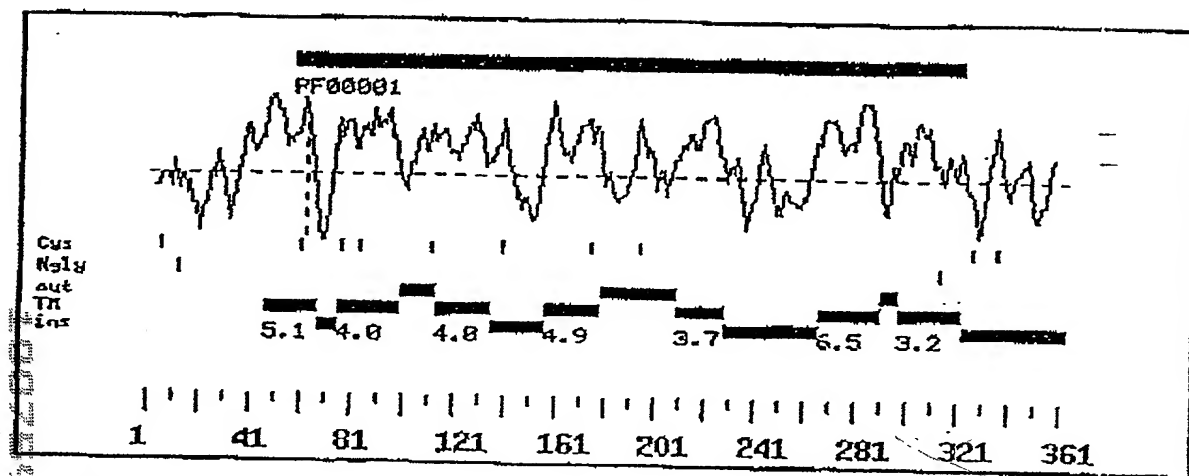


FIG 3

Analysis of Flh14273, (362 aa)



Flh14273, 1086 bases, 1825 checksum.
 MSPECARAAGDAPLRSLERANRTRPFFSDVKGDHRLVLAAVETTVLVLEFAVSLLGQNV
 ALVLVARRRRRGATACLVNLFPCADLLFISATPLVLAVRWTEAWLLGPVACHLLFYVMTL
 SGSVTILTAAVSLERMVCIIVHLQRGVVRGPGRRARAVLLALIWGYSAAALPLCVFFRVV
 FORLPGADQEISICTLIWPTIPGEISWDVSFVTLNFLVPLVIVISYSKILQITKASRR
 LTVSLAYSESHQIRVSQQDFRLFRFTLFLLMVSFFIMWSPITILLILIQNFQDLVIWP
 SLFFWVVAFTFANSALNPILYNMTLCRNEWKKIFCCFWFPEKGAILTDTSVKRNDLSIIS

FIG 4

Prosit Pattern Matches for Flh14273,

>PS00001|PDOC00001|ASN_GLYCOSYLATION N-glycosylation site.

Query: 21 NRTR 24

Query: 322 NMTL 325

>PS00004|PDOC00004|CAMP_PHOSPHO_SITE CAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 239 KRLT 242

>PS00005|PDOC00005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 237 SRK 239

Query: 350 SVK 352

>PS00006|PDOC00006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 256 SQDD 259

>PS00008|PDOC00008|MYRISTYL N-myristoylation site.

Query: 57 GNVCAL 62

Query: 72 CATACL 77

Query: 343 GAILTD 348

>PS00009|PDOC00009|AMIDATION Amidation site.

Query: 150 PGRR 153

>PS00022|PDOC00022|LEUCINE_ZIPPER Leucine zipper pattern.

Query: 106 LGPVACHLLFYVMTLSGVSITL 127

FIG 5

10075987.021300

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
46	66	out->ins	5.1
75	98	ins->out	4.0
113	134	out->ins	4.0
156	177	ins->out	4.9
209	227	out->ins	3.7
266	289	ins->out	6.5
297	321	out->ins	3.2

>Flh14273,

MSPECARAAGDAPLRSLQANRTRFFPFSDVKGDRHLVLAAVETTVLVLPFAVSLLGNVCL
ALVLVARRRRRGATACLVNLFCADLLFTISA_PVLAVRWTEAWLLGPVACHLLFYVMTL
SGSVTILTLAAVSLERMVCIVHLQRGVGPGRRARAVLLALIWGYSAVAALPLCVFFRVV
PQRLPGADQEISICTLIWPTIPGEISWDVSEVTLNPLVPGLVIVISYSKILQITKASRKR
LTVSLAYSESHQIRVSQODFRLFRTLFLLMVSFFIMWSPITITILLILIQNFQDLVIWP
SLFFWVVAFTFANSALNPILYNMTLCRNEWKKIFCCFWFPEKGAILEDTSVKRNDLSIIS

Transmembrane segments for presumed mature peptide

Start	End	Orient	Score
14	37	ins->out	4.0
52	73	out->ins	4.0
95	116	ins->out	4.9
148	166	out->ins	3.7
205	228	ins->out	6.5
236	260	out->ins	3.2

>Flh14273_mature

LVLVARRRRRGATACLVNLFCADLLFTISAIPVLAVRWTEAWLLGPVACHLLFYVMTLS
GSVTILTLAAVSLERMVCIVHLQRGVGPGRRARAVLLALIWGYSAVAALPLCVFFRVV
PQRLPGADQEISICTLIWPTIPGEISWDVSEVTLNPLVPGLVIVISYSKILQITKASRKR
LTVSLAYSESHQIRVSQODFRLFRTLFLLMVSFFIMWSPITITILLILIQNFQDLVIWPS
LFFWVVAFTFANSALNPILYNMTLCRNEWKKIFCCFWFPEKGAILEDTSVKRNDLSIISG

FIG 6

Input file 14273m; Output File 14273mtra
Sequence length 1560

TTGCCAAGCTCAGCGTAAGCCTCTTCCACTGCAATCTCACAGAAGGGGTTTCATGGAGTGCTTCACACCATCAGTGACCA

CTCCAGACTTGTCGGCTTTACCCGAATCTTCACAGCGGAGTCGATGACCTCTTGACAGCCACGAGCGCGCGGAGCTC

	M	S	P	E	C	A	Q	T	T	G		10
CGCCATCTTCCCGGACGCGTGGGCGGGCGCCCGGC	ATG	TCC	CCT	GAG	TGT	GCA	CAG	ACG	ACG	GGC		30
P G P S H T L D Q V N R T H F P F F S D												30
CCT GGT CCC TCG CAC ACC CTG GAC CAA GTC AAT CGC ACC CAC TTC CCT TTC TTC TCG GAT												90
V K G D H R L V L S V V E T T V L G L I												50
GTC AAG GGC GAC CAC CGG TTG GTG TTG AGC GTC GTG GAG ACC ACC GTT CTG GGA CTC ATC												150
F V V S L L G N V C A L V L V A R R R R												70
TTT GTC GTC TCA CTG CTG GGC AAC GTG TGT GCT CTA GTG CTG GTG GCG CGC CGT CGG CGC												210
R G A S A S L V L N L F C A D L L F T S												90
CGT GGG GCG TCA GCC AGC CTG GTG CTC AAC CTC TTC TGC GCG GAT TTG CTC TTC ACC AGC												270
A I P L V L V V R W T E A W L L G P V V												110
GCC ATC CCT CTA GTG CTC GTC GTG CGC TGG ACT GAG GCC TGG CTG TTG GGG CCC GTC GTC												330
C H L L F Y V M T M S G S V T I L T L A												130
TGC CAC CTG CTC TTC TAC GTG ATG ACA ATG AGC GGC AGC GTC ACG ATC CTC ACA CTG GCC												390
A V S L E R M V C I V R L R R G L S G P												150
GCG GTC AGC CTG GAG CGC ATG GTG TGC ATC GTG CGC CTC CGG CGC GGC TTG AGC GGC CCG												450
G R R T Q A A L L A F I W G Y S A L A A												170
GGG CGG CGG ACT CAG GCG GCA CTG CTG GCT TTC ATA TGG GGT TAC TCG GCG CTC GCC GCG												510
L P L Y I L F R V V P Q R L P G G D Q E												190
CTG CCC CTC TAC ATC TTG TTC CGC GTG GTC CCG CAG CGC CTT CCC GGC GGG GAC CAG GAA												570
I P I C T L D W P N R I G E I S W D V F												210
ATT CCG ATT TGC ACA TTG GAT TGG CCC AAC CGC ATA GGA GAA ATC TCA TGG GAT GTG TTT												630
F E T L N F L V P G L V I V I S Y S K I												230
TTT GAG ACT TTG AAC TTC CTG GTG CCG GGA CTG GTC ATT GTG ATC AGT TAC TCC AAA ATT												690
L Q I T K A S R K R L T L S L A Y S E S												250
TTA CAG ATC ACG AAA GCA TCG CGG AAG AGG CTT ACG CTG AGC TTG GCA TAC TCT GAG AGC												750
H Q I R V S Q Q D Y R L F R T L F L L M												270
CAC CAG ATC CGA GTG TCC CAA CAA GAC TAC CGA CTC TTC CGC ACG CTC TTC CTG CTC ATG												810
V S F F I M W S P I I I T I L L I L I Q												290
GTT TCC TTC TTC ATC ATG TGG AGT CCC ATC ATC ATC ACC ATC CTC CTC ATC TTG ATC CAA												870
N F R Q D L V I W P S L F F W V V A F T												310
AAC TTC CGG CAG GAC CTG GTC ATC TGG CCA TCC CTT TTC TTC TGG GTG GTG GCC TTC ACG												930
F A N S A L N P I L Y N M S L F R N E W												330
TTT GCC AAC TCT GCC CTA AAC CCC ATA CTG TAC AAC ATG TCG CTG TTC AGG AAC GAA TGG												990
R K I F C C F F F P E K G A I F T D S												350
AGG AAG ATT TTT TGC TGC TTC TTT TTT CCA GAG AAG GGA GCC ATT TTT ACA GAT ACG TCT												1050
V R R N D L S V I S S *												362
GTC AGG CGA AAT GAC TTG TCT GTT ATT TCC AGC TAA												1086

CTAGCCTCTGGTGCCAGGTGAACCACGGTGTGCATGTAAAGGGAGTTAACTTCAAGGAAAGCCCACCACTGCGCCCTTC

TTTAAAAATACCCGACTTCCAACAGCAGGCATCTACGGAGCCAGCAAATTAAGGAATGATCGCTCAGTATAAAAAATATT
TTTCCTTAAAAGAACTTTCTATGGGTCCTTTTGTGAACCTTTTAAAGTGTGTTGTAATATGATCTAGTTAATAAATT
TTTATTTATAACGTGTCCTACAAAAAAAAAAAAAAAAAAAAA

10075987.021302

Query: 14273m,

Scores for sequence family classification (score includes all domains):

Sequence	Description	Score	E-value	N
7tm_1	PF00001 7 transmembrane receptor (rhodopsin	118.8	1e-36	1

Parsed for domains:

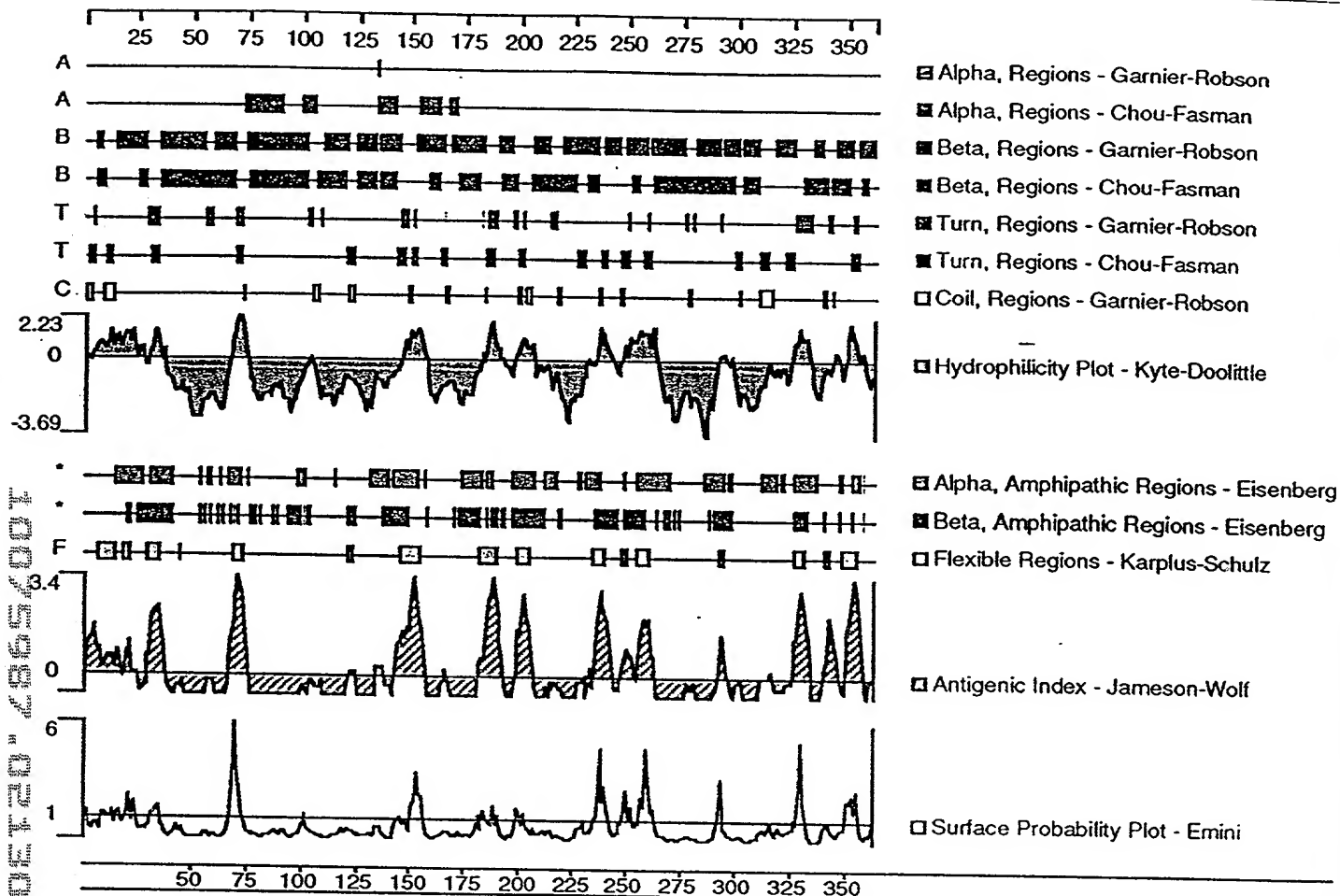
Sequence	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
7tm_1	1/1	57	321	..	1 259 []	118.8	1e-36

Alignments of top-scoring domains:

7tm_1: domain 1 of 1, from 57 to 321: score 118.8, E = 1e-36

```

*->GNLLVilvilrtkkkltptnifilNLAVADLLflltppwalyylvg
14273m, 57  GNVCALVLVAR-RRRRCASASLVNLFCADLLFSAIPVLVVR-WT 101
      gsdWpFGsalCklvtaldvnmmyaSilLLtaISiDRYLAiVhPlryrrr
      e W++G+++C+l+ ++++++ + il+L+a S++R + IV 1- +
14273m, 102 --EAWLLGPVVCHLLPYVMTKSGSVTILTLAAVSLERMVCIV-RLRRGLS 148
      rtsprrrA.kvwillvWvlalllslPpllfswvktveegngtlnvntvC1
      rr+++++++W ++l++lP ++++ v + ++g ++ +C+
14273m, 149 GP-GRRTQAALLAFIWGYSALAALPLYILFRVVPQRLPGGD--QEIPICT 195
      idfpeestaasvstwlrsyvlstlvgFl1Pl1lvilvcYtrIlrtlr....
      +d+p-- + ++s+ +++ ++ Fl+p lvi++ Y+ Il + + +++
14273m, 196 LDWPNRIG-----EISWDVFFETLNLFLVPGLVVISYSKILQITKsrk 239
      .....kaaktllvvvvvFvlCWlPyfivllldt1
      + + + +++++ + ++++ ++ +tl++++v F++ W P i++ll +
14273m, 240 rltlslayseshqirvsqgcyRLFRTLFLLMVSFFIMWSPIIITILLILI 289
      c.lsiimsstCelervlptallv1wLayvNsc1NPi1Y<-*
      ++ + + P +++++ + ++++Ns+1NPi+Y
14273m, 290 QnFR-----QDLVIWPSLFFWVVAFTFANSALNPILY 321
  
```



Prosite Pattern Matches for 14273m,

>PS00001|PDOC00001|ASN_GLYCOSYLATION N-glycosylation site.

Query: 21 NRTH 24
Query: 322 NMSL 325

>PS00002|PDOC00002|GLYCOSAMINOGLYCAN Glycosaminoglycan attachment site.

RU Additional rules:
RU There must be at least two acidic amino acids (Glu or Asp) from -2 to
RU -4 relative to the serine.

Query: 148 SGPG 151

>PS00004|PDOC00004|CAMP_PHOSPHO_SITE CAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 239 KRLT 242

>PS00005|PDOC00005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 237 SRX 239

Query: 350 SVR 352

>PS00006|PDOC00006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 40 SVVE 43

Query: 256 SQOD 259

>PS00008|PDOC00008|MYRISTYL N-myristoylation site.

Query: 57 GNVCAL 62

Query: 72 GASASL 77

Query: 343 GAIFTD 348

>PS00009|PDOC00009|AMIDATION Amidation site.

Query: 150 PGRR 153

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
46	66	out->ins	5.1
77	98	ins->out	3.3
113	134	out->ins	3.8
156	177	ins->out	4.5
209	227	out->ins	2.4
266	289	ins->out	6.5
297	321	out->ins	3.2

>14273m,

MSPECAQTGPGPSHTLDQVNRTHFPFFSDVKGDHRLVLSVVETTTLGLIFVVSLLGNVC
 ALVLVARRRRRGASASLVNLFCADLLFTSAIPVLVVRWTEAWLLGPVVCHLLFYVMTM
 SGSVTILTLAAVSLERMVCIVRLRGLSGPGRRTQAALLAFIWGYSALAALPLYILFRVV
 PQRPGGDQEIPICTLDWPNRIGEISWDVFFETLNFVPLVIVISYSKILQITKASRRK
 LTLSESLAYSESHQIRVSQQDYRLFRTEFLLMVSFFIMWSPIITILLILIQNFRQDLVIWP
 LFFVWVAPTTFANSALNPILYNMSLPRNEWRKIFCCFFPEKGAIFTDTSVRRNDLSVIS
 S

Transmembrane segments for presumed mature peptide

Start	End	Orient	Score
16	37	ins->out	3.3
52	73	out->ins	3.8
95	116	ins->out	4.5
148	166	out->ins	2.4
205	228	ins->out	6.5
236	260	out->ins	3.2

>14273m, _mature

LVLVARRRRRGASASLVNLFCADLLFTSAIPVLVVRWTEAWLLGPVVCHLLFYVMTM
 GSVTILTLAAVSLERMVCIVRLRGLSGPGRRTQAALLAFIWGYSALAALPLYILFRVV
 PQRPGGDQEIPICTLDWPNRIGEISWDVFFETLNFVPLVIVISYSKILQITKASRRK
 LTLSESLAYSESHQIRVSQQDYRLFRTEFLLMVSFFIMWSPIITILLILIQNFRQDLVIWP
 LFFVWVAPTTFANSALNPILYNMSLPRNEWRKIFCCFFPEKGAIFTDTSVRRNDLSVIS
 S